**UNIVERSIDAD AUTÓNOMA DE MADRID**

**ESCUELA POLITÉCNICA SUPERIOR**

**Gráfico de barras

Descripción generada automáticamente con confianza media**

**Grado en Ingeniería Informática**

**TRABAJO FIN DE GRADO**

**Plantilla Memoria TFG**

**Autor: Álvaro Delgado Sancho**

**Tutor: Daniel Perdices Burrero**

**Junio 2024**

### Agradecimientos

Me gustaría mostrar mi agradecimiento, en primer lugar, a mi tutor Daniel Perdices por su gran ayuda y disposición a lo largo de todo el desarrollo de este trabajo, así como a la totalidad del departamento de HPCN (High Processing Computing and Networking) de la Escuela Politécnica Superior (EPS) de la Universidad Autónoma de Madrid (UAM).

Además, dar las gracias a todo el equipo docente de la EPS por acompañarme durante estos 4 años de grado poniendo su tiempo a disposición de una formación de calidad, y por su alta exigencia y motivación que me han ayudado a ser un mejor estudiante y profesional.

Por otro lado, agradecer a mis padres, Francisco Javier y María de los Milagros, y hermanos, Francisco Javier y Verónica, junto al resto de familiares y amigos por el afecto y apoyo incondicional mostrado durante todo este proceso de aprendizaje universitario. Finalmente, agradecer a mi pareja Paula por su cariño y motivación.

En resumen, agradecer a todas las personas que hayan podido ayudar o tener un impacto, con mayor o menor intensidad, en la realización de este trabajo fin de grado y que han hecho posible que este trabajo haya llegado a buen término.

### Resumen

Resumen

**Palabras clave**

Palabras clave

### Abstract

Abstract

**Keywords**

Keywords

### Índice

[Agradecimientos III](#_Toc167956739)

[Resumen V](#_Toc167956740)

[Abstract VII](#_Toc167956741)

[Índice IX](#_Toc167956742)

[Índice de figuras XII](#_Toc167956743)

[Índice de tablas XII](#_Toc167956744)

[Índice de cuadros XII](#_Toc167956745)

[1 Introducción 1](#_Toc167956746)

[1.1. Motivación 1](#_Toc167956747)

[1.2. Objetivos 1](#_Toc167956748)

[2 Estado del Arte 2](#_Toc167956749)

[2.1. Motivación 2](#_Toc167956751)

[2.2. Objetivos 2](#_Toc167956752)

[3 Desarrollo 4](#_Toc167956753)

[3.1. Subsección 4](#_Toc167956756)

[3.2. Subsección 4](#_Toc167956757)

[3.3. Subsección 4](#_Toc167956758)

[3.3.1. Subsubsección 4](#_Toc167956762)

[4 Experimentos 6](#_Toc167956766)

[4.1. Subsección 6](#_Toc167956768)

[4.2. Subsección 6](#_Toc167956769)

[4.3. Subsección 6](#_Toc167956770)

[4.3.1. Subsubsección 6](#_Toc167956773)

[5 Resultados 7](#_Toc167956774)

[5.1. Subsección 7](#_Toc167956776)

[5.2. Subsección 7](#_Toc167956777)

[5.3. Subsección 7](#_Toc167956778)

[5.3.1. Subsubsección 7](#_Toc167956781)

[6 Conclusiones 8](#_Toc167956782)

[6.1. Subsección 8](#_Toc167956784)

[6.2. Subsección 8](#_Toc167956785)

[6.3. Subsección 8](#_Toc167956786)

[6.3.1. Subsubsección 8](#_Toc167956789)

[7 Trabajo Futuro 9](#_Toc167956790)

[7.1. Subsección 9](#_Toc167956792)

[7.2. Subsección 9](#_Toc167956793)

[7.3. Subsección 9](#_Toc167956794)

[7.3.1. Subsubsección 9](#_Toc167956797)

[Bibliografía 9](#_Toc167956798)

[Apéndices 11](#_Toc167956799)

[Apéndice A 13](#_Toc167956800)

### Índice de figuras

X.X Figura 1 XX

### Índice de tablas

X.X Tabla 1 XX

### Índice de cuadros

X.X Cuadro 1 XX

# Introducción

En los últimos años, investigadores y administradores de sistemas han tomado una gran relevancia en el ámbito de la monitorización de servicios de red con el objetivo de detectar comportamientos extraños y generar alertas para dichas situaciones. Esto sumado a la creciente capacidad y carga de servidores de empresas modernas y CPD (Centro de Procesado de Datos), ha hecho de este proceso algo bastante complejo y laborioso, debido a la variedad en el comportamiento de los servidores y subredes de organizaciones y grandes empresas. El objetivo principal de esta labor es la de encontrar soluciones que se adapten al comportamiento concreto de los servidores o segmentos de red a analizar. La gran cantidad de información y datos que poseen las series temporales, así como el limitado tiempo de los investigadores y administradores, dificulta su manejo y nos sugiere poner la vista en soluciones basadas en aprendizaje automático que puedan hacer frente al procesado masivo de datos.

La monitorización de segmentos de red es una herramienta muy útil para caracterizar las series temporales del tráfico de redes y poder así agruparlas respecto a un comportamiento similar. La agrupación de estos servicios nos permite ver un comportamiento común sobre el que poder detectar anomalías o situaciones de cambio relevantes en la dinámica del segmento de red.

La detección de estos cambios o anomalías suponen una gran ventaja para las organizaciones y empresas de cara a poder distribuir sus recursos de manera eficiente y lógica, y tener capacidad de previsión acerca de posibles subidas o bajadas de carga en sus servidores. La incapacidad para prever este tipo de situaciones podría suponer la pérdida de disponibilidad de servicios o el derroche de estos en una red infrautilizada, lo que se traduce en problemas económicos para la empresa.

Para la caracterización de estas series temporales se suelen utilizar métodos de clustering, que de manera automática agrupan los datos en el número de clústeres elegido teniendo en cuenta posibles relaciones entre los datos de su mismo clúster y diferencias más o menos significativas frente a datos de otros clústeres.

Para el proceso de agrupación de las series temporales, teniendo en cuenta la gran cantidad de información que poseen estas, sería necesario hacer uso de alguna técnica de reducción de componentes como PCA, t-SNE o UMAP. Posteriormente, se utilizarían métodos de clustering como k-means o DBSCAN para generar los resultados en base a la reducción de componentes previamente realizada.

Esta reducción de componentes nos facilita la visualización de los datos para su interpretación posterior, cuyo concepto surge de la idea de que la representación de una secuencia o grupo de datos puede basarse en un subconjunto más pequeño de variables respecto al inicial, manteniendo siempre medida de la varianza entre los datos. La varianza es la métrica que ayuda a distribuir los datos sobre los distintos clústeres y que mide la diferencia o variabilidad de los datos respecto al resto, por lo que al reducir el número de componentes se debe mantener dicha variabilidad para que la distribución de los datos se mantenga correctamente, que es justamente la idea que implementan los métodos de visualización como PCA (técnica que hemos utilizado).

Una vez están agrupados los datos en los distintos clústeres, la idea es poder categorizar las relaciones o patrones que puedan existir entre las series temporales pertenecientes a un mismo clúster, de manera que podamos encontrar información útil para poder utilizarla posteriormente en detección de anomalías o similares.

Introducción

#### Motivación

Motivación

#### Objetivos

Objetivos:

* Objetivo1
* Objetivo2

# Estado del Arte

Introducción



#### Motivación

Motivación

#### Objetivos

Objetivos:

* Objetivo1
* Objetivo2

# Desarrollo

Estado del arte



#### Subsección

Subsección

#### Subsección

Subsección

#### Subsección



##### Subsubsección

Gráfico de barras

Descripción generada automáticamente con confianza baja

**Figura 2.1:** Logo EPS



# Experimentos

Estado del arte



#### Subsección

Subsección

#### Subsección

Subsección

#### Subsección



##### Subsubsección

# Resultados

Estado del arte



#### Subsección

Subsección

#### Subsección

Subsección

#### Subsección



##### Subsubsección

# Conclusiones

Estado del arte



#### Subsección

Subsección

#### Subsección

Subsección

#### Subsección



##### Subsubsección

# Trabajo Futuro

Estado del arte



#### Subsección

Subsección

#### Subsección

Subsección

#### Subsección



##### Subsubsección

### Bibliografía

|  |  |
| --- | --- |
| [1] | Fuente1 |
| [2] | Fuente2 |

### Apéndices

A

### Apéndice A

Apéndice A